PAPER Air Quality Index Forecasting via Deep Dictionary Learning

SUMMARY Air quality index (AQI) is a non-dimensional index for the description of air quality, and is widely used in air quality management schemes. A novel method for Air Quality Index Forecasting based on Deep Dictionary Learning (AQIF-DDL) and machine vision is proposed in this paper. A sky image is used as the input of the method, and the output is the forecasted AQI value. The deep dictionary learning is employed to automatically extract the sky image features and achieve the AQI forecasting. The idea of learning deeper dictionary levels stemmed from the deep learning is also included to increase the forecasting accuracy and stability. The proposed AQIF-DDL is compared with other deep learning based methods, such as deep belief network, stacked autoencoder and convolutional neural network. The experimental results indicate that the proposed method leads to good performance on AQI forecasting.

key words: air quality index forecasting, deep dictionary learning, greedy learning, representation learning paradigm

1. Introduction

With the rapid development of industrialization and urbanization, air pollution is increasing. Each year, more than 4 million people die early because of outdoor air pollution [1]. Especially in developing countries, serious air pollution occurred in certain seasons. Li et al. [2] introduced that ambient air pollution in China poses a multifaceted health threat to outdoor physical activity. The suggestion that health authorities in China must address the critical dilemma of how to protect and encourage the active population is also proposed by Li et al. W. James et al. [3] proved that air quality in southern California is associated with statistically and clinically significant improvements in childhood lung-function growth. Therefore, the air quality is associated with our daily life.

Air quality index (AQI) is a non-dimensional index for the description of air quality, and is based on the level of several atmospheric pollutants, namely sulfur dioxide (SO₂), nitrogen dioxide (NO₂), suspended particulates smaller than 10 μ m in aerodynamic diameter (PM10), suspended particulates smaller than 2.5 μ m in aerodynamic diameter (PM2.5), carbon monoxide (CO), and ozone (O₃). Different countries have their own air quality indices, corresponding to different national air quality standards, and the measurement is at the monitoring stations. Notably, AQI is essential indicator to evaluate air quality, the higher the number, the greater the

Manuscript revised January 14, 2020.

[†]The author is with College of mathematics physics and information engineering, Jiaxing University, Jiaxing 314000, China.

a) E-mail: chenbin@zjxu.edu.cn

Bin CHEN^{†a)}, Member

health risks and the need for preventive measures. The public mainly focus on measuring the AQI conveniently and by themselves [4], [5]. Kang Z. et al. [6] proposed a BP neural network based algorithm to predict the AQI via six atmospheric pollutants. Wang et al. [7] introduced a hybrid forecasting model for PM10 and SO₂ daily concentrations. The support vector machine (SVM) is used to achieve the forecasting, a 2-year dataset of daily PM10 and SO₂ is also needed. But the forecasting is only focused on the PM10 and SO₂, the measurement of dataset is difficult too. Wu Q. et al. [8] proposed an optimal-hybrid model for daily air quality index prediction considering air pollutant factors, but the inputs of the model are the six atmospheric pollutants.

Most existed AQI forecasting methods lack real-time and universal property in different countries. The public neither can measure the AOI by themselves as the complexity of the measurement. For these issues, a new Air Quality Index Forecasting method based on Deep Dictionary Learning (AQIF-DDL) and machine vision is proposed in this paper. The input of the method is just an image which only contains the sky region, and the output is the air quality index of the image. The public can capture the image conveniently even via their phones, and the AQI is forecasted immediately. Dictionary learning is a kind of representation learning paradigm, focuses on learning "basis" and "features" by matrix factorization, the current popularity dictionary learning owes to K-SVD [9], [10]. Deep learning is another kind of representation learning paradigm, focuses on extracting features via learning "weights" and "features" in a greedy layer by layer method [11]–[13]. In order to increase the forecasting accuracy and stability, the proposed AQIF-DDL combined the two representation learning paradigms, learned multi-level deep dictionaries.

2. Related Works

According to different emphasis on the prior knowledge, most existed AQI forecasting methods can be categorized into four groups [8]: the deterministic model based, the statistical model based, the artificial intelligent model based, the hybrid model based. i) The deterministic model based methods utilize the atmospheric physics data and model to calculate the AQI. However, the atmospheric conditions are complex, the calculation model is relatively deterministic, which suffers the difficult application problem [14]. ii) The statistical model based methods achieve the prediction

Manuscript received November 7, 2019.

Manuscript publicized February 20, 2020.

DOI: 10.1587/transinf.2019EDP7296

via the mathematical statistical algorithms and criteria air pollutants, but the prediction accuracy is relatively low [15]. iii) The artificial intelligent model based methods also suffer the low prediction accuracy problem as the limitation of database size and computational ability [16]-[20]. iv) The hybrid model based methods focus on enhancing the time series via the signal decomposition technology and combined model, the sub-sequences are obtained by these methods [21]–[24]. But the hybrid model based methods are lack of air pollution prediction, the prediction accuracy is also non-stationary. Wang et al. [25] proposed an AQI forecasting method via support vector machine (SVM), moments and machine version, but the prediction accuracy is unsatisfactory. The main reason of low prediction accuracy in [25] is the feature extraction problem, only the color moments and wavelet features are utilized to achieve the AQI forecasting. Additionally, the convolutional neural network based methods are not suitable as the amount limitation of dataset.

In this study, a novel deep dictionary learning based air quality index forecasting method is developed. The input of the proposed method is a sky image, and the output is the corresponding AQI value. The deep dictionary learning is employed to automatically extract the sky image features and achieve the AQI forecasting. Dictionary learning is a kind of representation learning paradigm, focuses on learning a basis for representation in early studies. The basis of dictionary learning is to solve the optimization problem:

$$\min_{D,X} \|Y - DX\|_F^2 \tag{1}$$

Where *Y* is the training data, *X* is the loading coefficients and *D* is a dictionary to be learned. In sparse representation problem, *X* is sparse enough to be solved, K-SVD [9], [10] is one of the most well know methods to achieve the optimization, the common format is:

$$\min_{D,X} ||Y - DX||_F^2 \quad \text{subject to } ||X||_0 \le \tau \tag{2}$$

The l_0 -norm is defined on the vectorized version of X. The solution of l_0 -norm is a NP-hard [26], [27] problem, K-SVD utilizes the greedy (sub-optimal) orthogonal matching pursuit (OMP) [28] to achieve the solution. In the dictionary learning step, K-SVD develops an efficient method to estimate and update the atoms one at a time. The discriminative K-SVD [29] or LC-KSVD [30] is proposed to learn discriminative sparse representation, which had a good results on face recognition. The proposed AQIF-DDL method in this paper extends the state-of-the-art in AQI forecasting. The public can achieve the AQI forecasting by themselves, the prediction process is convenient and the accuracy is satisfactory.

This paper is organized as follows. The first section is the introduction of air quality index and the forecasting. The second section describes the main existed AQI forecasting methods and dictionary learning. The third section introduces the deep dictionary learning and the proposed AQIF-DDL method. The fourth section is the experimental results part, and the last section is the conclusion part.

3. Methods

In this section, we describe the deep dictionary learning and the proposed AQIF-DDL method. The deep dictionary learning is a kind of greedy layer-wise learning, which learns the latent representation of data by learning multilevel dictionaries. The idea of learning deeper levels of dictionaries is employed in deep dictionary learning, which is the main success reason of deep learning.

3.1 Deep Dictionary Learning

In order to explain the concept of deep dictionary learning clearly, a two-layer deep dictionary learning is introduced, then we extend it to a multi-level dictionary [31]. Figure 1 shows the schematic diagram of dictionary learning:

Where the *Y* is the data, *X* is the loading coefficients and D_1 is a dictionary to be learned. The dictionary learning follows the representation paradigm:

$$Y = D_1 X \tag{3}$$

The analysis K-SVD is used to find the solution of (3), but the result is without redundancy, which cannot extract the features from the data (*Y*). The result gained by analysis K-SVD is only suitable for inverse problems. Hence, the two-layer deep dictionary learning was developed [31] to fix that problem. The schematic diagram of two-layer deep dictionary learning is shown in Fig. 2. The representation paradigm of (3) can be rewritten as:

$$Y = D_1 D_2 X \tag{4}$$

Learning two-layer dictionaries (a tri-linear problem) is really different from learning a single dictionary (a bi-linear



Fig. 1 Schematic diagram of dictionary learning



Fig. 2 Schematic diagram of two-layer dictionary

learning problem) [32], [33]. The over-fitting problem occurred in deep learning as the amount limitation of training data is also existed in learning two-layer dictionaries. However, the greedy learning has been successfully utilized to overcome these issues [34]–[37]. The schematic diagram of two-layer greedy layer-wise dictionary learning is shown in Fig. 3.

The Eq. (4) is rewritten as:

$$Y = D_1 \varphi(D_2 X_2) \tag{5}$$

where φ is an activation function, it can be linear or nonlinear. One layer is learned at a time in the employed greedy dictionary learning. Hence, the D_1 and X_1 in Fig. 3 are solved by:

$$\min_{D_1, X_1} \|Y - D_1 X_1\|_F^2 \tag{6}$$

The solving of (6) has been achieved by the alternating minimization in [38]. Then, the D_1 and X_1 are alternating learned by:

$$X_1 \leftarrow \min_{X_1} \|Y - D_1 X_1\|_F^2$$
 (7)

$$D_1 \leftarrow \min_{D_1} \|Y - D_1 X_1\|_F^2 \tag{8}$$



Fig. 3 Schematic diagram of greedy dictionary learning

The Eqs. (7) and (8) are belonged to simple least square problems, it is easy to be calculated. For the next layer, the X_1 , D_2 and X_2 have the relation of $X_1 = \varphi(D_2X_2)$. The relational expression also can be rewritten as $\varphi^{-1}(X_1) = D_2X_2$, it is a single layer dictionary learning. The solution of X_2 is dense, it also can be solved by:

$$\min_{D_2, X_2} \|\varphi^{-1}(X_1) - D_2 X_2\|_F^2 \tag{9}$$

The layers in the greedy dictionary learning are solved via alternating learning till the penultimate layer. In the last layer, the relation is $\varphi^{-1}(X_{N-1}) = D_N X$. In order to obtain features, the regulation by l_1 -norm is employed:

$$\min_{D_N,X} \|\varphi^{-1}(X_{N-1}) - D_N X\|_F^2 + \lambda \|X\|_1$$
(10)

where λ is a coefficient, the D_N and X are alternating calculated by the Iterative Soft Thresholding Algorithm (ISTA) [39]:

$$X \leftarrow \min_{Y} \|\varphi^{-1}(X_{N-1}) - D_N X\|_F^2 + \lambda \|X\|_1$$
(11)

$$D_N \leftarrow \min_{D_N} \|\varphi^{-1}(X_{N-1}) - D_N X\|_F^2$$
 (12)

The deep dictionary learning is achieved via the greedy learning and alternating minimization, the features are automatically extracted via the deep dictionary structure, and it can be used in the AQI forecasting task with a further improvement in structure.

3.2 Air Quality Index Forecasting via Deep Dictionary Learning (AQIF-DDL)

The flow chart of the proposed AQIF-DDL method is shown in Fig. 4, it consists of N layers with different dictionary separately $(D_i, i = 1, 2, ..., N)$. The input of the proposed



Fig. 4 Flow chart of the proposed AQIF-DDL with N layers. The dictionary atoms shown in the flow chart are 3-dimensional, which correspond to the color image RGB channels.

method is a color image only contains the sky region, the output is the forecasted AQI value (V_{AQI}). The trained deep dictionary structure is used to automatically extract the air quality features, and a classifier (SVM) is also utilized to achieve the AQI value forecasting. The dictionary atoms used in the deep dictionary structure are 3-dimensional, which correspond to the color image RGB channels.

Usually the image size is larger than the size of dictionary atom. Hence, we utilized the patch extraction [40] to gain the patches with 5 pixels overlap, the extracted patch has the same size as dictionary atom. Then, the forecasted AQI value of the input image is calculated by:

$$V_{AQI} = \frac{1}{m} \sum_{z=1}^{m} v_{AQI}(z)$$
(13)

Where $v_{AQI}(z)$ is the forecasted AQI value of patch z, m is the total number of patches, V_{AQI} is the AQI value of the input image, it is the mean of all patches. In order to extract the air quality features, we trained the deep dictionary one layer a time via greedy dictionary learning, the training and testing algorithms are shown below.

Algorithm 1 : Training AQIF-DDL

- (a) Initialize: $D_i, i = 1, 2, ...N, l$ is the layer number
- (b) The first layer: repeat until convergence

$$X_1 \leftarrow \min_{X_1} \left\| Y - D_1 X_1 \right\|_F^2$$
$$D_1 \leftarrow \min_{D_1} \left\| Y - D_1 X_1 \right\|_F^2$$

(c) The second layer to penultimate layer: repeat until convergence

$$X_{l} \leftarrow \min_{X_{l}} \left\| \varphi^{-1}(X_{l-1}) - D_{l} X_{l} \right\|_{F}^{2}$$
$$D_{l} \leftarrow \min_{D_{l}} \left\| \varphi^{-1}(X_{l-1}) - D_{l} X_{l} \right\|_{F}^{2}$$

(d) The last layer: repeat until convergence

$$X \leftarrow \min_{X} \left\| \varphi^{-1}(X_{N-1}) - D_{N}X \right\|_{F}^{2} + \lambda \left\| X \right\|_{1}$$
$$D_{N} \leftarrow \min_{D_{Y}} \left\| \varphi^{-1}(X_{N-1}) - D_{N}X \right\|_{F}^{2}$$

(e) Train the classifier (SVM).

Algorithm 2 : Testing AQIF-DDL

- (a) Calculate the output of the first layer: $X_{1,test} \leftarrow \min_{X_{1,test}} \left\| Y_{test} - D_1 X_{1,test} \right\|_F^2$
- (b) The second layer to penultimate layer:

$$X_{l,test} \leftarrow \min_{X_{l,test}} \left\| \varphi^{-1}(X_{l-1}) - D_1 X_{l,test} \right\|_F^2$$

(c) The last layer:

$$X_{test} \leftarrow \min_{X_{test}} \left\| \varphi^{-1}(X_{N-1}) - D_N X_{test} \right\|_F^2 + \lambda \left\| X_{test} \right\|$$

(d) Calculate the V_{AOI} via classifier and Eq.(13).

4. Results

In order to evaluate the forecasting performance of the proposed method, the RMSE, MAE and MAPE are employed. The definition of the evaluation indices are shown below:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(V_{AQI,i} - \hat{V}_{AQI,i} \right)^2}$$
(14)

$$MAE = \frac{1}{n} \sum_{i=1}^{n} \left| V_{AQI,i} - \hat{V}_{AQI,i} \right|$$
(15)

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{V_{AQI,i} - \hat{V}_{AQI,i}}{\hat{V}_{AQI,i}} \right|$$
(16)

The $V_{AQI,i}$ is forecasted AQI value of image *i*, the $\tilde{V}_{AQI,i}$ is the actual data. It can be known that the smaller the criterion, the better forecasting performance is. The experimental results are compared with the state-of-the-art algorithms, such as deep belief network (DBN) [41], stacked autoencoder (SAE) [42], D-KSVD [29], LC-KSVD [30] and convolutional neural network [43]. We utilized the linear activation function for the dictionary learning and SVM.

4.1 Datasets

In the experiment, the dataset consists of 3000 sky images gained from 500 days in Beijing with the AQI value range [21, 420]. The image size is $100 \times 100 \times 3$, which only contains the sky region. We utilized 2500 sky images to train the multi-level deep dictionary, and 500 images as the testing set. Figure 5 shows part of the dataset with different AQI values, it can be observed that the sky region in the image gradually turns yellow as the AQI value increases. When the AQI value is larger than 300, the air quality level is moderately polluted. People with breathing or heart problems will experience reduced endurance in activities. These individuals and elders should remain indoors and restrict activities in these situation.

The size of dictionary atom is usually smaller than the input image, we utilized the patch extraction to gain the





Fig. 6 Schematic diagram of the image patch extraction

 Table 1
 The effect of different layers on testing dataset.

| Evaluation | Shallow(1 layer) | 2 layer | 3 layer | 4 layer | 5 layer |
|------------|---------------------|------------|------------|------------|------------|
| RMSE | 13.15 | 12.40 | 9.97 | 9.94 | 9.91 |
| MAE | 8.93 | 7.62 | 6.71 | 6.70 | 6.69 |
| MAPE | 0.152 | 0.141 | 0.098 | 0.096 | 0.095 |

patches with 5 pixels overlap. Hence, the dictionary atom and the input have the same size. The schematic diagram of the patch extraction is shown in Fig. 6, where the size of the extracted patch is $20 \times 20 \times 3$. The total number of patches is 36 in a sky image, where m = 36 in Eq. (13).

4.2 Effect of Different Layers

The proposed AQIF-DDL consists of multi-level deep dictionary, we first analyzed the effect of increasing the number of layers. One layer dictionary is usually used in the dictionary learning based works, it is a kind of shallow structure. We learned 1200, 2400-1200, 2400-1600-1200, 2400-1600-1200-1200, 2400-1600-1200-1200-1200 atoms for the shallow dictionary, 2 layer, 3 layer, 4 layer and 5 layer separately. The λ used in Eq. (11) is 0.12 in all layers, Table 1 shows the effect of different layers on testing set. In Table 1, column 6 has the best performance, we observed that the forecasting results have a positive correlation with the number of dictionary layers. The main reason of these phenomenon is that the representations learned from multilevel dictionaries have a better AQI forecasting accuracy than the single level. However, the multi-level dictionaries training time also should be mentioned. The training time of 4 layer is 10 times that of 3 layer, but the improvement of AQI forecasting accuracy is limited. Hence, we utilized 3 layer structure in the rest of the experiment as the comprehensive consideration of training speed and forecasting accuracy.

4.3 Effect of Dictionary Atom Amount in Each Layer

The amount of the dictionary atom in AQIF-DDL also affects the AQI forecasting accuracy, we analyzed the effect of atom amount via layer by layer. Figure 7 shows the AQI forecasting results with different atom amounts, the *k* is a variable used to represent the value of the dictionary atom amount, such as 2400-*k*-1200. The increasing step of the atom amount is 100, the size of dictionary atom is $20 \times 20 \times 3$. It can be observed that the evaluation indices decrease as the *k* increases in all layers, and the rate of decrease is gradually slowed down after a certain value. In that situation, the features used for the AQI forecasting had been efficiently extracted via the deep dictionary structure, the increase of *k* has little effect on the forecasting results.

From the comparison of different layers, we also observed that the last layer in the deep dictionary structure has the maximum effect on the AQI forecasting. Especially from the first point comparison of MAE in Fig. 7 (a1), (b1) and (c1), the MAE is 8.29 (2400-1600-400 dictionary atom structure), which is larger than the other structures (1600-1600-1200 dictionary atom structure and 2400-800-1200 dictionary atom structure).

4.4 Effect of Dictionary Atom Size

The dictionary atom size is another significant parameter in the dictionary learning, a suitable value leads to a good performance for the dictionary learning based method. We analvzed the effects of different atom size on AOI forecasting in this section. The atom size is from $3 \times 3 \times 3$ to $30 \times 30 \times 3$, and the dictionary atom amount is 2400-1600-1200. Table 2 shows the effect of different dictionary atom size on testing dataset, the image patch size is same as the atom size. The overlap pixel used in patch extraction is an integer which is a quarter of atom size, for instance, the overlap pixel of $15 \times 15 \times 3$ is 4. From the comparison of different dictionary atom size, we observed that the small size do not have a good performance on the AQI forecasting. The main reason is that the AQI features extracted by these atoms can not effectively characterize the difference between sky images. Hence, the $20 \times 20 \times 3$ is the suitable atom size, which has the best performance on AQI forecasting. At the last row of Table 2, we also presented the two-dimensional atom, extracted from the gray image, but the results are not ideal. The missing of color information may be the main reason.

4.5 Comparison with Other Deep Learning Methods

In order to demonstrate the advantages of the proposed AQIF-DDL method, five other deep learning based methods are taken as the comparison, which is deep belief network (DBN) [41], stacked autoencoder (SAE) [42], D-KSVD [29], LC-KSVD [30] and convolutional neural network (CNN) [43]. The dictionary atom structure of AQIF-DDL used in this section is 2400-1600-1200, the atom size

0.12 0.1

14PE

MAE-testing data
 MAE-training data

MAE of training and testing data via AQIF-DDL(k-1600-1200 Dictionary atom)

MAE

MAE

MAE





MAPE-testing d MAPE-training dat

RMSE

1000 1200 1400 1800 2000 800 Dictionary atom amount (k) (c1)MAE of 2400-1600-k dictionary atom





(c3)RMSE of 2400-1600-k dictionary atom

The AQI forecasting results with different atom amounts. Fig. 7

Table 3 Comparison with other methods on testing dataset.

| | | | - |
|--------------|-------|------|-------|
| Method | RMSE | MAE | MAPE |
| SAE [41] | 11.58 | 7.61 | 0.152 |
| DBN [42] | 11.47 | 7.48 | 0.149 |
| D-KSVD [29] | 12.81 | 8.75 | 0.181 |
| LC-KSVD [30] | 12.96 | 8.96 | 0.19 |
| AQIF-DDL | 9.97 | 6.71 | 0.098 |
| CNN [43] | 9.95 | 6.78 | 0.101 |

is $20 \times 20 \times 3$. Figure 8 shows part of the dictionary atoms in the first layer.

Similar to deep dictionary learning, the SAE and DBN also utilize a three-layer structure, the parameters used in D-KSVD, LC-KSVD and CNN are similar to [29], [30] and [43] separately. Table 3 shows the AQI forecasting results with different methods on testing dataset. It can be observed that the proposed AQIF-DDL has the best performance on

Table 2 The effect of different dictionary atom size on testing dataset.

| Atom size | RMSE | MAE | MAPE |
|------------------------------|-------|-------|-------|
| $3 \times 3 \times 3$ | 13.58 | 8.97 | 0.161 |
| $5 \times 5 \times 3$ | 13.02 | 8.75 | 0.152 |
| $7 \times 7 \times 3$ | 12.54 | 8.03 | 0.146 |
| 11×11×3 | 11.12 | 7.65 | 0.131 |
| 13×13×3 | 10.85 | 7.48 | 0.126 |
| $15 \times 15 \times 3$ | 10.26 | 7.05 | 0.108 |
| $17 \times 17 \times 3$ | 10.25 | 7.03 | 0.102 |
| $20 \times 20 \times 3$ | 9.97 | 6.71 | 0.098 |
| $23 \times 23 \times 3$ | 9.98 | 6.72 | 0.098 |
| $27 \times 27 \times 3$ | 10.01 | 6.71 | 0.099 |
| $30 \times 30 \times 3$ | 10.02 | 6.72 | 0.100 |
| 20×20 (gray image) | 18.47 | 12.24 | 0.456 |

2400



Fig. 8 The first 625 atoms in the first layer of the proposed AQIF-DDL method

semi-supervised learning methods (except the CNN based method [43], which has the smallest RMSE), the MAE is almost 25% smaller than the LC-KSVD [30] based method. The SAE [41] and DBN [42] based methods have the similar accuracy on evaluation indices, but the MAPE is almost 34% larger than our method. The D-KSVD [29] based method has the penultimate AQI forecasting results in all comparison methods.

The CNN based method [43] is a kind of supervised learning which needs a large training dataset. However, the training dataset (only contain 2500 sky images) is limited in the AQI forecasting task, the CNN based method [43] suffers the overfitting problem. Hence, the semi-supervised learning methods have a better performance on these applications as the relatively loose data requirements. The MAE gained from our method is 6.71, which is better than the most existed atmospheric pollutants based methods [8]. The AQI forecasting results via the proposed AQIF-DDL are accurate and stable, the public can achieve the forecasting conveniently and immediately. It is not only an AQI forecasting method to promote urban public health, but also crucial for sustainable development of environment under the negative impact of air pollution.

5. Conclusions

In this research, we proposed a novel method for air quality index forecasting based on deep dictionary learning and machine vision. The input of the proposed AQIF-DDL is just a sky image, the output is the forecasted AQI value. A deep dictionary structure is developed to automatically extract the sky image features and achieve the AQI forecasting. The idea of learning deeper dictionary levels is stemmed from the deep learning, and the greedy learning is employed to achieve the deep dictionary training. The image patch extraction is also employed to reduce the dictionary atom size and implement the AQI forecasting. In the experimental part, we analyzed the effect of layer amount, dictionary atom in each layer and dictionary atom size. We observed that 3 layer structure has an ideal performance as the comprehensive consideration of training speed and forecasting accuracy. We also found the suitable atom amount and size via the experiments. From the comparison with other semisupervised learning methods, it can be observed that the results gained by our method is accurate, the proposed AOIF-DDL has the best performance on AQI forecasting task. The public can achieve the AQI forecasting conveniently and immediately via the proposed method, it may be a new measure to promote urban public health. Future work includes further improving AQI forecasting accuracy, increasing the collection of AQI image data, and applying it to other air quality forecasting task.

Acknowledgments

This research was funded by Science and technology planning project of Jiaxing, China, Grant number 2018AY11001; Zhejiang Provincial Natural Science Foundation of China under Grants, Grant number LQ20F020026, LQ18F020007 and LY18F020021.

References

- X. Li, L. Jin, and H. Kan, "Air pollution: A global problem needs local fixes," Nature, vol.570, pp.437–439, 2019.
- [2] F. Li, Y. Liu, J. Lü, L. Liang, and P. Harmer, "Ambient air pollution in China poses a multifaceted health threat to outdoor physical activity," Journal of epidemiology and community health, vol.69, no.3, pp.201–204, 2015.
- [3] W.J. Gauderman, R. Urman, E. Avol, K. Berhane, R. McConnell, E. Rappaport, R. Chang, F. Lurmann, and F. Gilliland, "Association of improved air quality with lung development in children," New England Journal of Medicine, vol.372, pp.905–913, 2015.
- [4] S. Radisic, K.B. Newbold, J. Eyles, and A. Williams, "Factors influencing health behaviors in response to the air quality health index: a cross-sectional study in Hamilton, Canada," Environmental Health Review, vol.59, no.1, pp.17–29, 2016.
- [5] J. Zavala, J.D. Krug, S.H. Warren, Q.T. Krantz, C. King, J. McKee, S.H. Gavett, M. Lewandowski, W.A. Lonneman, T.E. Kleindienst, M.J. Meier, M. Higuchi, M.I. Gilmour, and D.M. DeMarini, "Evaluation of an air quality health index for predicting the mutagenicity of simulated atmospheres," Environmental science & technology, vol.52, pp.3045–3053, 2018.
- [6] Z. Kang and Z. Qu, "Application of BP neural network optimized by genetic simulated annealing algorithm to prediction of air quality index in Lanzhou," 2017 2nd IEEE International Conference on Computational Intelligence and Applications (ICCIA), pp.155–160, 2017.
- [7] P. Wang, Y. Liu, Z. Qin, and G. Zhang, "A novel hybrid forecasting model for PM10 and SO₂ daily concentrations," Science of The Total Environment, vol.505, pp.1202–1212, 2015.
- [8] Q. Wu and H. Lin, "A novel optimal-hybrid model for daily air quality index prediction considering air pollutant factors," Science of The Total Environment, vol.683, pp.808–821, 2019.
- [9] R. Rubinstein, A.M. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," Proc, IEEE, vol.98, no.6, pp.1045–1057, 2010.
- [10] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm

for designing overcomplete dictionaries for sparse representation," IEEE Trans. Signal Process., vol.54, no.11, pp.4311–4322, 2006.

- [11] Y. Ding and Y. Xue, "A Deep Learning Approach to Writer Identification Using Inertial Sensor Data of Air-Handwriting," IEICE Trans. Inf. & Syst., vol.E102-D, no.10, pp.2059–2063, 2019.
- [12] Y. Xie, R. Liang, Z. Liang, and L. Zhao, "Attention-Based Dense LSTM for Speech Emotion Recognition," IEICE Trans. Inf. & Syst., vol.E102-D, no.7, pp.1426–1429, 2019.
- [13] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol.521, pp.436–444, 2015.
- [14] K. Vijayaraghavan, S. Cho, R. Morris, D. Spink, J. Jung, R. Pauls, and K. Duffett, "Photochemical model evaluation of the ground-level ozone impacts on ambient air quality and vegetation health in the Alberta oil sands region: using present and future emission scenarios," Atmospheric environment, vol.141, pp.209–218, 2016.
- [15] O. Taylan, "Modelling and analysis of ozone concentration by artificial intelligent techniques for estimating air quality," Atmospheric environment, vol.150, pp.356–365, 2017.
- [16] D. Voukantsis, K. Karatzas, J. Kukkonen, T. Räsänen, A. Karppinen, and M. Kolehmainen, "Intercomparison of air quality data using principal component analysis, and forecasting of PM10 and PM2.5 concentrations using artificial neural networks, in Thessaloniki and Helsinki," Science of the Total Environment, vol.409, no.7, pp.1266–1276, 2011.
- [17] S. Saad, A. Andrew, A. Shakaff, A.R.M. Saad, A.M. Yusof, and A. Zakaria, "Classifying sources influencing indoor air quality (IAQ) using artificial neural network (ANN)," Sensors, vol.15, no.5, pp.11665–11684, 2015.
- [18] J. He, Y. Yu, Y. Xie, H. Mao, L. Wu, N. Liu, and S. Zhao, "Numerical model-based artificial neural network model and its application for quantifying impact factors of urban air quality," Water, Air, & Soil Pollution., vol.227, p.235, 2016.
- [19] N.H.A. Rahman, M.H. Lee, Suhartono, and M.T. Latif, "Artificial neural networks and fuzzy time series forecasting: an application to air quality," Quality & Quantity, vol.49, pp.2633–2647, 2015.
- [20] J.C. Patni and H.K. Sharma, "Air Quality Prediction using Artificial Neural Networks," 2019 International Conference on Automation, Computational and Technology Management (ICACTM), pp.568–572, 2019.
- [21] L.-S. Chang, A. Cho, H. Park, K. Nam, D. Kim, J.-H. Hong, and C.-K. Song, "Human-model hybrid Korean air quality forecasting system," Journal of the Air & Waste Management Association, vol.66, pp.896–911, 2016.
- [22] Q. Di, P. Koutrakis, and J. Schwartz, "A hybrid prediction model for PM2.5 mass and components using a chemical transport model and land use regression," Atmospheric environment, vol.131, pp.390–399, 2016.
- [23] M. Niu, Y. Wang, S. Sun, and Y. Li, "A novel hybrid decomposition-and-ensemble model based on CEEMD and GWO for shortterm PM2.5 concentration forecasting," Atmospheric environment, vol.134, pp.168–180, 2016.
- [24] Y. Hao and C. Tian, "The study and application of a novel hybrid system for air quality early-warning," Applied Soft Computing, vol.74, pp.729–746, 2019.
- [25] W. Wang, W. Shen, B. Chen, R. Zhu, and Y. Sun, "Air Quality Index Forecasting Based on SVM and Moments," 2018 5th International Conference on Systems and Informatics (ICSAI), pp.851–855, 2018.
- [26] B.K. Natarajan, "Sparse approximate solutions to linear systems," SIAM journal on computing, vol.24, no.2, pp.227–234, 1995.
- [27] S. Jokar and M.E. Pfetsch, "Exact and approximate sparse solutions of underdetermined linear equations," SIAM Journal on Scientific Computing, vol.31, no.1, pp.23–44, 2008.
- [28] Y.C. Pati, R. Rezaiifar, and P.S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," Proc. 27th Asilomar conference on signals, systems and computers, pp.40–44, 1993.

- [29] Q. Zhang and B. Li, "Discriminative K-SVD for dictionary learning in face recognition," 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.2691–2698, 2010.
- [30] Z. Jiang, Z. Lin, and L.S. Davis, "Label consistent K-SVD: Learning a discriminative dictionary for recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol.35, no.11, pp.2651–2664, 2013.
- [31] S. Tariyal, A. Majumdar, R. Singh, and M. Vatsa, "Deep dictionary learning," IEEE Access, vol.4, pp.10096–10109, 2016.
- [32] P. Jain, P. Netrapalli, and S. Sanghavi, "Low-rank matrix completion using alternating minimization," Proc. forty-fifth annual ACM symposium on Theory of computing, pp.665–674, 2013.
- [33] A. Agarwal, A. Anandkumar, P. Jain, and P. Netrapalli, "Learning sparsely used overcomplete dictionaries via alternating minimization," SIAM Journal on Optimization, vol.26, no.4, pp.2775–2799, 2016.
- [34] A. Majumdar and R. Ward, "Robust greedy deep dictionary learning for ECG arrhythmia classification," 2017 International Joint Conference on Neural Networks (IJCNN), pp.4400–4407, 2017.
- [35] P.B. Nair, A. Choudhury, and A.J. Keane, "Some greedy learning algorithms for sparse regression and classification with mercer kernels," Journal of Machine Learning Research, vol.3, pp.781–801, 2002.
- [36] S. Harmeling and C.K.I. Williams, "Greedy learning of binary latent trees," IEEE Trans. Pattern Anal. Mach. Intell., vol.33, no.6, pp.1087–1097, 2010.
- [37] E. Belilovsky, M. Eickenberg, and E. Oyallon, "Decoupled Greedy Learning of CNNs," arXiv preprint arXiv:1901.08164, pp.1–15, 2019.
- [38] A. Makhzani and B. Frey, "K-sparse autoencoders," arXiv preprint arXiv:1312.5663, pp.1–9, 2013.
- [39] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences, vol.57, no.11, pp.1413–1457, 2004.
- [40] B. Chen, Y. Chen, Z. Shao, T. Tong, and L. Luo "Blood vessel enhancement via multi-dictionary and sparse coding: Application to retinal vessel enhancing," Neurocomputing, vol.200, pp.110–117, 2016.
- [41] M. Jiang, Y. Liang, X. Feng, X. Fan, Z. Pei, Y. Xue, and R. Guan, "Text classification based on deep belief network and softmax regression," Neural Computing and Applications, vol.29, pp.61–70, 2018.
- [42] Y. Bai, Y. Li, B. Zeng, C. Li, and J. Zhang, "Hourly PM2.5 concentration forecast using stacked autoencoder model with emphasis on seasonality," Journal of Cleaner Production, vol.224, pp.739–750, 2019.
- [43] K. Rujirakul and C. So-In, "Histogram equalized deep PCA with ELM classification for expressive face recognition," 2018 International Workshop on Advanced Image Technology (IWAIT), pp.1–4, 2018.



Bin Chen received his Ph.D. degree in computer science and technology from Southeast University, China, in 2017. He is currently a lecturer at the College of mathematics physics and information engineering, Jiaxing University. His research interests include deep learning, image analysis and pattern recognition.